

What is Claimed is:

1. A method for removing a vector unit base sequence from a DNA base sequence which is obtained as a result of performing a cloning process by integrating an object DNA fragment into a vector and includes the vector unit base sequence as a part of a base sequence of the vector and the object DNA fragment, comprising the steps of:
- generating a retrieval base sequence as a retrieval key for use in retrieving the vector unit base sequence from the DNA base sequence based on the vector, a restriction enzyme used for cleaving the vector for the cloning process, and a restriction enzyme used for obtaining the object DNA fragment;
- specifying the vector unit base sequence using the retrieval key; and
- removing the specified vector unit base sequence to specify the object DNA fragment.
2. The method according to Claim 1, wherein said DNA base sequence is obtained as an output from a sequencer for determining the DNA base sequence.
3. The method according to Claim 1, wherein

said retrieval key comprises a forward retrieval key and a backward retrieval key for respectively identifying areas before and after the object DNA fragment in the DNA base sequence; and

5 said forward and backward retrieval keys indicate base sequences corresponding to restriction enzyme sites including parts of the vector cleaved by a restriction enzyme for the cloning process and ends of the object DNA fragment.

10

4. The method according to Claim 3, wherein
base sequences of the forward and backward retrieval keys are generated by base sequence data of the vector entered in a vector data base, data of a multiple cloning site in the vector, and data of a restriction enzyme site in the multiple cloning site.

15

5. The method according to Claim 3 further comprising the steps of:

20 performing homology retrieval on condition that a similarity value indicating a matching rate between the retrieval base sequence and the DNA base sequence is equal to or larger than a predetermined value in retrieval using the retrieval key for the DNA base
25 sequence; and

obtaining a candidate for a base sequence at a junction between the vector in the DNA base sequence and the object DNA fragment according to a result of the homology retrieval.

5

6. The method according to Claim 5 further comprising the steps of:

generating a second forward retrieval key by adding to the forward retrieval key a portion that should be existing before the multiple cloning site of the vector;

performing a second homology retrieval on condition that a second similarity value indicating a matching rate between a base sequence corresponding to the second forward retrieval key and a base sequence including a base sequence at a junction of the DNA base sequence is equal to or larger than a predetermined value; and

obtaining as a vector unit candidate for the vector unit base sequence an area specified as a result of the second homology retrieval and an area before the specified area.

7. The method according to Claim 5 further comprising the steps of:

25

generating a second backward retrieval key by adding to the backward retrieval key a portion that should be existing after the multiple cloning site of the vector;

5 performing a second homology retrieval on condition that a second similarity value indicating a matching rate between a base sequence corresponding to the second backward retrieval key and a base sequence including a base sequence at a junction of
10 the DNA base sequence is equal to or larger than a predetermined value; and

obtaining as a vector unit candidate for the vector unit base sequence an area specified as a result of the second homology retrieval and an area
15 after the specified area.

8. The method according to Claim 6, wherein
said vector unit candidate is removed from the DNA
base sequence when the number of the areas specified
20 by the second homology retrieval is one.

9. The method according to Claim 7, wherein
said vector unit candidate is removed from the DNA
base sequence when the number of the areas specified
25 by the second homology retrieval is one.

10. The method according to Claim 5 further comprising:

generating a second forward retrieval key by adding to the forward retrieval key a portion that should be existing before the multiple cloning site of the vector;

generating a second backward retrieval key by adding to the backward retrieval key a portion that should be existing after the multiple cloning site of the vector;

performing a second homology retrieval on condition that a second similarity value indicating a matching rate between a base sequence corresponding to the second forward retrieval key and a base sequence including a base sequence at a junction of the DNA base sequence is equal to or larger than a predetermined value, and a third similarity value indicating a matching rate between a base sequence corresponding to the second backward retrieval key and a base sequence including a base sequence at a junction of the DNA base sequence is equal to or larger than a predetermined value;

obtaining as a forward vector unit candidate for the vector unit base sequence a forward area specified as a result of the second homology retrieval and an

area before the forward area; and

obtaining as a backward vector unit candidate for the vector unit base sequence a backward area specified as a result of the second homology retrieval and an area after the backward area.

11. The method according to Claim 10, wherein said forward vector unit candidate and said backward vector unit candidate are removed from the DNA base sequence when there is only one candidate respectively for the specified forward and backward vector units, and the specified forward and backward vector units do not overlap each other.

12. A device for removing a vector unit base sequence from a DNA base sequence which is obtained as a result of performing a cloning process by integrating an object DNA fragment into a vector and includes the vector unit base sequence as a part of a base of the vector and the object DNA fragment, comprising:

first means for generating a base sequence as a retrieval key for use in retrieving the vector unit base sequence from the DNA base sequence based on the vector, a first restriction enzyme used for cleaving the vector for the cloning process, and a second

restriction enzyme used for obtaining the object DNA fragment;

second means for specifying the vector unit base sequence using the retrieval key; and

5 third means for removing the specified vector unit base sequence to specify the object DNA fragment.

13. The device according to Claim 12 further comprising:

10 vector list storage means for storing a vector list; and

restriction enzyme list storage means for storing a restriction enzyme list, wherein

15 said vector is specified in the vector list, and said first and second restriction enzymes are specified in the restriction enzyme list.

14. The device according to Claim 13 further comprising:

20 display means, wherein

said vector is specified in the vector list displayed on said display means; and

25 at least one of said first and second restriction enzymes is specified in the restriction enzyme list displayed on said display means.

15. The device according to Claim 12 further comprising:

program storage means for storing at least one of:

a program for generating the retrieval key by
controlling said first means;

a program for specifying the vector unit base
sequence by controlling said second means; and

a program for removing the vector unit base
sequence by controlling said third means.

16. The device according to Claim 12, wherein

said second means specifies using the retrieval
key a junction between the vector unit base sequence
and the object DNA fragment; and

said third means specifies the object DNA fragment
by removing the junction and a portion outside the
junction from the DNA base sequence.

17. The device according to Claim 16, wherein

said second means specifies as the junction a
portion in the DNA sequence in which a number of bases
matching a base sequence of the retrieval key is equal
to or larger than a predetermined value.

18. The device according to Claim 12, wherein

said second means specifies using the retrieval key a first junction and a second junction between the vector unit base sequence and the object DNA fragment; and

5 said third means specifies the object DNA fragment by removing from the DNA base sequence, the first junction and a portion outside the first junction, and the second junction and a portion outside the second junction .

10

19. The device according to Claim 12, wherein

 said retrieval key includes a base sequence corresponding to an end portion of the object DNA fragment and a base sequence corresponding to an end portion of the vector unit base sequence, and
15 specifies a candidate for a junction between the vector unit base sequence and the object DNA fragment.

20. The device according to Claim 19, wherein
20 a second retrieval key indicating a base sequence longer than said retrieval key is generated, and the junction is specified in the candidates for the junction using the second retrieval key.

25 21. The device according to Claim 20, wherein

said object DNA fragment is specified by removing the junction and a portion outside the junction from the DNA base sequence.

- 5 22. A storage medium embodying a program for performing by a computer a function of removing a vector unit base sequence from a DNA base sequence which is obtained as a result of performing a cloning process by integrating an object DNA fragment into a
- 10 vector and includes the vector unit base sequence as a part of a base sequence of the vector and the object DNA fragment, said program realizing the steps of:
- generating a retrieval base sequence as a retrieval key for use in retrieving the vector unit
- 15 base sequence from the DNA base sequence based on the vector, a restriction enzyme used for cleaving the vector for the cloning process, and a restriction enzyme used for obtaining the object DNA fragment;
- specifying the vector unit base sequence using the
- 20 retrieval key; and
- removing the specified vector unit base sequence to specify the object DNA fragment.

Add
A2